ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ
INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND MANAGEMENT



УДК 004.421.2 DOI 10.12737/20222

Метод кластеризации в системе оценки результатов работы экспертов по проверке экзаменов в форме ЕГЭ*

Т. А. Щербинин[1], М. В. Гранков[2]**
[1, 2]Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация

Clustering method for quality verification of experts' Unified State Examination checking ***

T. A. Shcherbinin[1], M. V. Grankov[2]**
[1, 2] Don State Technical University, Rostov-on-Don, Russian Federation

Цель данного исследования — рассмотрение возможности использования метода нечеткой кластеризации в автоматизированной системе оценивания эффективности работы экспертов по проверке результатов государственной итоговой аттестации в форме ЕГЭ. Представлены различные показатели результатов работы экспертов и их расчетные формулы. Параметры разбиты на следующие группы: базовые, статистические оценки базовых, нормированные и качественные. Применение метода нечеткой кластеризации с использованием взвешенной метрики Евклида показано на примере оценивания работы экспертов предметной комиссии по математике (профильный уровень). Эксперименты проводились с учетом и без учета шумов с последующими выводами о влиянии, которое оказывают шумы на результаты. Также по результатам экспериментов были представлены таблицы и графики с разбиением экспертов на кластеры.

The possibilities of using the fuzzy clustering method in the automated system of evaluation working efficiency of the experts' validation status on the assessment of the state final certification in the form of the Unified State Exam (USE) are considered. Various parameters of experts' evaluation and their calculation formula are presented. The parameters are divided into the following groups: basic statistical parameters, advanced statistical parameters, normalized parameters, and quality parameters. The application of the fuzzy clustering method using a weighted Euclidean metric is illustrated by the case of the assessment of the experts' work in the subject commission on mathematics (profession-oriented level). The experiments are performed with and without noises with the subsequent conclusions of the impact of noises on the results. Besides, the experimental results are presented in tables and graphs with partitioning of experts into clusters.

Ключевые слова: нечеткая логика, кластерный анализ, ЕГЭ, оценивание, шумы в данных, терм-множества.

Keywords: fuzzy logic, cluster analysis, Unified State Exam, evaluation, noise in the data, term-sets.

Введение. С 2009 года в России единый государственный экзамен (ЕГЭ) является единственной формой государственной итоговой аттестации обучающихся в организациях общего среднего образования и одновременно вступительных испытаний в вузы.

По большинству предметов задания делятся на три группы в зависимости от их сложности. Первая группа — задания области *A*, самые простые. В них предлагается выбрать правильный ответ из предложенных вариантов. Вторая группа — задания области В. Они сложнее, предполагают самостоятельное написание краткого ответа. Ответы на задания групп *A* и *B* вносятся в бланк и проверяются с помощью компьютера. Третья группа — задания области *C*, самые сложные. Они требуют развернутых ответов. Например, предлагается написать эссе, привести решение задачи, дать обоснованный ответ на вопрос и т. д. Тексты ответов вносятся в бланк и проверяются экспертами региональной экспертной комиссии. Следует отметить, что нередки случаи некорректного выставления баллов экспертами при проверке заданий области *C*. В связи с этим конечный балл, полученный экзаменуемым, зависит не только от правильности выполненных им заданий, но и от компетентности проверяющих [1].

Таким образом, важной задачей является формирование квалифицированной экспертной комиссии, проверяющей задания области *C*.

В данной работе предложен метод, позволяющий ответственным за формирование предметных комиссий обосновано разделять экспертов на кластеры в зависимости от значений показателей их работы. Такая кластеризация

* Работа выполнена в рамках инициативной НИР.
** E-mail: sherbinin5@mail.ru, mv_2@mail.ru
*** The research is done within the frame of the independent R&D.

позволит руководителям принимать решения об уровне квалификации экспертов, необходимости переподготовки некоторых из них или исключении из состава комиссии.

Ошибки при проверке экзаменационных работ. Процедура проверки экспертом экзаменационных работ области *С* регламентируется инструкциями Федерального института педагогических измерений (ФИПИ), которые содержат подробные схемы (алгоритмы), определяющие число баллов, выставляемых за выполнение заданий каждого варианта. Экспертам предоставляются отсканированные копии работ.

Анализ результатов работы экспертов позволяет утверждать, что при выполнении проверки возможны ошибки трех типов:

1) эксперт не обнаруживает имевшееся решение задачи и, следовательно, не оценивает его;
2) эксперт неверно понимает ход решения задачи;
3) эксперт неверно понимает инструкцию ФИПИ по процедуре оценивания решения задания.

Ошибки 1-го типа будем называть ошибками идентификации решения [2]. Ошибки 2-го и 3-го типов будем называть ошибками оценивания.

Для сокращения влияния ошибок экспертов на оценку заданий области С предусмотрена процедура проверки решения каждого задания двумя экспертами. Процедуру проверки задания вторым экспертом будем называть второй проверкой. Если суммы баллов разных экспертов различаются несущественно, то учитывается большая из них.

Перед проведением экзамена для каждого предмета ФИПИ устанавливает критерий несущественности различия баллов. Например, при проверке заданий по математике несущественным будет отличие в два балла. Если расхождение больше, то данное задание отправляется на проверку третьему эксперту. Процедуру проверки задания третьим экспертом будем называть третьей проверкой. Третью проверку выполняет опытный эксперт высокой квалификации, выставленный им балл считается окончательным.

Основная идея метода определения корректности работы экспертов. Рассмотрим оценивание одного и того же задания разными экспертами. Примем следующие допущения.

— Оценивая задания, эксперты действуют независимо друг от друга.
— Если суммы баллов в двух случаях совпадают, то можно считать малой вероятность одновременного принятия двумя экспертами ошибочного решения.
— Если суммы баллов в двух проверках существенно различаются, то вероятность ошибки третьего эксперта мала в силу его высокой квалификации.
— Если эксперт допустил ошибку первого типа (не обнаружил задачу), то его оценка за эту задачу принимается равной 0.
— Определение корректности результатов работы экспертов основывается на значениях оценок, удельных относительно количества проверенных заданий, с учетом параметров случайных процессов.

Будем оценивать различия в баллах, выставленных за задание с номером *k* при второй ($\Delta x2_i^k$) и третьей ($\Delta x3_i^k$) проверке, по формулам:

$$\Delta x2_i^k = x2_j^k - x2_i^k , \quad (1)$$

где $x2_i^k$ — оценка, выставленная за *k*-е задание *i*-м экспертом; $x2_j^k$ — оценка, выставленная за *k*-е задание *j*-м экспертом, в паре с которым был *i*-й эксперт.

Соответственно,

$$\Delta x3_i^k = x3_j^k - x3_i^k , \quad (2)$$

где $x3_i^k$ — оценка, выставленная за *k*-е задание *i*-м экспертом; $x3_j^k$ — оценка, выставленная за *k*-е задание *j*-м экспертом, осуществляющим третью проверку.

Корректность работы экспертов в комиссии может быть оценена рядом показателей. Разделим их на четыре группы.

1. Базовые выявляются при первичном анализе результатов проверки комиссией области *С*.

2. Статистические формируются на основе оценки базовых показателей работы экспертов.

3. Нормированные строятся для базовых и соответствующих им статистических оценок по множествам экспертов [3].

4. Качественные представляют собой оценку параметров на основании результатов выполнения алгоритма кластеризации и нечеткого логического вывода.

Список рассматриваемых базовых показателей результатов работы экспертов представлен в табл. 1.

Таблица 1

Базовые показатели результатов работы экспертов

№	Показатель	Сокращение
1	Количество проверенных заданий	*ОбщЗдн*
2	Количество заданий, по которым оценка (сумма баллов) расходится с результатом второй проверки	*2ПрКор*
3	Количество заданий, для которых была проведена третья проверка	*3ПрОбщ*
4	Количество заданий, по которым оценка (сумма баллов) расходится с результатом третьей проверки	*3ПрКор*
5	Сумма абсолютных значений расхождений с суммой баллов третьей проверки	*3ПрАбсКор*

Показатель *ОбщЗдн* отражает объем работы, выполненной экспертом. Показатель *2ПрКор* отражает количество расхождений с результатами второй проверки, но не дает возможности понять, кто из двух экспертов допустил ошибку. Показатель *3ПрОбщ* отражает количество ошибок эксперта, выявленных при третьей проверке. Показатель *3ПрАбсКор* определяет, насколько существенными были ошибки эксперта.

В табл. 2 представлены статистические оценки базовых показателей работы экспертов.

Таблица 2

Статистические оценки показателей работы экспертов

№	Показатель	Сокращение
1	Показатель слаженности работы экспертов	*СлажГрп*
2	Среднее значение расхождений при второй проверке ($\Delta x2_i$)	*2ПрСрКор*
3	Среднеквадратическое отклонение балла расхождений при второй проверке	*СКО2ПрСрКор*
4	Частота возникновения у эксперта корректировок при вторых проверках	*2ПрЧастКор*
5	Среднее значение балла корректировок при третьей проверке ($\Delta x3_i$)	*3ПрСрКор*
6	Среднеквадратическое отклонение корректировок при третьей проверке	*СКО3ПрСрКор*
7	Частота возникновения третьих проверок у эксперта	*3ПрЧаст*
8	Частота возникновения корректировок эксперта при третьих проверках	*3ПрЧастКор*
9	Удельное значение корректируемого балла для эксперта	*3ПрУдВелКор*

Исходя из принятых допущений, качество работы эксперта можно оценить с помощью показателя *СлажГрп*, который вычисляется по формуле:

$$СлажГрп = \frac{n_c^{грп}}{n^{грп}}, \qquad (3)$$

где $n^{грп}$ — общее число задач экзаменующихся, решение которых было оценено в группе эксперта; $n_c^{грп}$ — число задач, оценка (сумма баллов) которых совпала с результатами второй проверки.

Показатели *2ПрСрКор* и *3ПрСрКор* могут использоваться при принятии решения о наличии соответствующей тенденции у оцениваемого эксперта.

Показатель *2ПрСрКор* находится по формуле:

$$2ПрСрКор = \frac{\sum_{k=1}^{t2} \Delta x2_i^k}{t2}, \qquad (4)$$

где $t2$ — количество расхождений при второй проверке у i-го эксперта; $\Delta x2_i^k$ — расхождение выставленного i-м экспертом балла в k-м задании по формуле (1).

Показатель *3ПрСрКор* находится по формуле:

$$3ПрСрКор = \frac{\sum_{k=1}^{t3} \Delta x3_i^k}{t3}, \qquad (5)$$

где $t3$ — количество корректировок эксперта при третьей проверке; $\Delta x3_i^k$ — корректировка выставленного i-м экспертом балла в k-м задании по формуле (2).

Показатели *2ПрЧастКор*, *3ПрЧаст* и *3ПрЧастКор* характеризуют частоту несовпадений оценок экспертов при вторых проверках, частоту возникновения третьей проверки и частоту корректировки при третьей проверке соответственно. Данные показатели находятся относительно общего числа проверенных заданий. Для определения этих показателей использовались формулы:

$$2ПрЧастКор = \frac{2ПрКор}{ОбщЗдн}; \tag{6}$$

$$3ПрЧаст = \frac{3ПрОбщ}{ОбщЗдн}; \tag{7}$$

$$3ПрЧастКор = \frac{3ПрКор}{ОбщЗдн}. \tag{8}$$

Показатель *3ПрУдВелКор* отражает количество ошибочно выставленных баллов, приходящееся на одно проверенное экспертом задание. Данный параметр находится по формуле:

$$3ПрУдВелКор = \frac{\sum_{k=1}^{t3}\left|\Delta x3_i^k\right|}{ОбщЗдн}, \tag{9}$$

где $t3$ — количество корректировок эксперта при третьей проверке; $\Delta x3_i^k$ — отличие от выставленного i-м экспертом балла в k-м задании — в соответствии с формулой (2).

Для построения системы сравнения качества работы экспертов рационально использовать нормированные значения оценок — будем называть их индикаторами (имеют приставку «*Инд*»). Индикаторы соответствуют количественным показателям оценки эксперта. Например, *Инд3ПрСрКор* рассчитан для параметра *3ПрСрКор*. Значения индикаторов находятся в интервале [0;1]: 1 соответствует наилучшему значению параметра, 0 — наихудшему. Рассматриваются 7 нормированных параметров:

— *Инд2ПрСрКор,*
— *ИндСКО2ПрСрКор,*
— *Инд2ПрЧастКор,*
— *Инд3ПрСрКор,*
— *ИндСКО3ПрСрКор,*
— *Инд3ПрЧастКор,*
— *Инд3ПрУдВелКор.*

Показатель $Инд3ПрЧастКор_i$ для i-го эксперта находится по формуле:

$$Инд3ПрЧастКор_i = \frac{\max(3ПрЧастКор) - 3ПрЧастКор_i}{\max(3ПрЧастКор) - \min(3ПрЧастКор)}, \tag{10}$$

где $\max(3ПрЧастКор)$ — максимальное значение показателя *3ПрЧастКор* среди всех членов комиссии; $\min(3ПрЧастКор)$ — минимальное значение показателя *3ПрЧастКор* среди всех членов комиссии.

Аналогично равенству (10) (и при соответствующей замене показателей $3ПрЧастКор_i$ и *3ПрЧастКор*) находятся величины $Инд2ПрЧастКор_i$, $Инд3ПрУдВелКор_i$, $ИндСКО2ПрСрКор_i$ и $ИндСКО3ПрСрКор_i$.

Итак, завышение или занижение баллов при оценивании заданий является отрицательной характеристикой эксперта. Предпочтительно отсутствие любого из названных недостатков. Поэтому показатели $Инд2ПрСрКор_i$ и $Инд3ПрСрКор_i$ для i-го эксперта рассчитываются по формулам:

$$Инд2ПрСрКор_i = \frac{\max(|2ПрСрКор|) - |2ПрСрКор_i|}{\max(|2ПрСрКор|) - \min(|2ПрСрКор|)}, \tag{11}$$

$$Инд3ПрСрКор_i = \frac{\max(|3ПрСрКор|) - |3ПрСрКор_i|}{\max(|3ПрСрКор|) - \min(|3ПрСрКор|)}, \tag{12}$$

где $\max(|2ПрСрКор|)$, $\max(|3ПрСрКор|)$ — максимальные абсолютные значения соответствующих показателей среди всех членов комиссии; $\min(|2ПрСрКор|)$, $\min(|3ПрСрКор|)$ — минимальные абсолютные значения соответствующих показателей среди всех членов комиссии.

При обосновании принятия решений по оцениванию результатов работы экспертов была применена теория нечетких множеств. Данная теория предложена Л. Заде и позволяет описывать нечеткие понятия и знания, оперировать этими знаниями и делать нечеткие выводы. В нечеткой логике значения любой величины представляются не чис-

лами, а словами естественного языка и называются термами [4, 5]. При оценивании качества работы экспертов количественные показатели дополняются качественными оценками, выраженными базовыми и расширенными терм-множествами. Для качественных оценок значений показателей были использованы три основных терма базового терм-множества: «низкое» (низкое качество проверки), «среднее» (среднее качество проверки), «высокое» (высокое качество проверки). Также были введены четыре терма, относящиеся к расширенному терм-множеству: «скорее низкое, чем среднее», «скорее среднее, чем низкое», «скорее среднее, чем высокое» и «скорее высокое, чем среднее». Термы из расширенного терм-множества позволяют классифицировать экспертов, которые явно не относятся ни к одной из основных групп. Использовались следующие качественные параметры эксперта (имеют приставку «*Оц*»):
— *Оц2ПрСрКор*,
— *Оц2ПрЧастКор*,
— *Оц3ПрСрКор*,
— *Оц3ПрЧастКор*.

Метод нечеткой кластеризации. В качестве метода кластеризации экспертов по результатам их работы был использован алгоритм *Fuzzy C-Means* (*FCM*) [6, 7]. Этот алгоритм чаще всего применяется для нечеткого группирования объектов. Он формирует группы, прототипы которых представляются точками в пространстве параметров кластерных объектов. Для выполнения алгоритма группировки объектов с множеством параметров A необходимо выбрать подмножество параметров объектов $B \subset A$, по значениям которых будет реализована кластеризация, определить подходящую метрику, количество групп c, степень нечеткости n, параметр ε в критерии остановки алгоритма, а также случайным образом инициировать матрицу принадлежности объектов к той или иной группе $U^{(0)} \in Z_i$ (i — количество групп) и вектор прототипов $V^{(0)}$ [8].

Были проведены эксперименты по обработке результатов проверки работ различных экзаменационных комиссий [1, 2, 10]. С помощью анализа показателей результатов работы экспертов с учетом принятых допущений установлено, что качество работы экспертов наиболее информативно отражают следующие показатели: *Инд3ПрЧастКор*, *Инд3ПрУдВелКор, Инд3ПрСрКор* и *ИндСКО3ПрСрКор*. Показатель *Инд3ПрЧастКор* представляет нормированную частоту появления ошибок у экспертов при оценивании результатов выполнения задания. *Инд3ПрУдВелКор* отражает существенность ошибок эксперта при оценивании. *Инд3ПрСрКор* оценивает наличие тенденции к занижению или завышению оценок при выставлении баллов. *ИндСКО3ПрСрКор* демонстрирует стабильность в принятии экспертом решения. Для наглядности было выбрано двумерное пространство параметров, построенное по индикаторам, оценивающим частоту корректировки при третьей проверке (параметр y) и удельное значение корректируемого балла оценивания при третьей проверке (параметр x).

В рассматриваемом двумерном пространстве координаты точки — это оценки результатов работы экспертов. Расстояние между точками может быть измерено различными метриками, которые являются формулами (правилами) для определения расстояния между любыми двумя точками заданного пространства. При выборе метрики были рассмотрены несколько возможных вариантов: расстояние Чебышева, Евклидово расстояние, квадрат Евклидова расстояния, «взвешенное» Евклидово расстояние, расстояние городских кварталов (Манхэттенское расстояние).

Для рассматриваемых параметров экспертов приемлемые результаты кластеризации были получены с применением метрики взвешенного расстояния Евклида.

Расстояние ($dist_i$) от точки до центра i-го кластера находится по формуле:

$$dist_i = \sqrt{k_1(x_i - x)^2 + k_2(y_i - y)^2}\ , \qquad (13)$$

где x_i, y_i — значения координат центра i-го кластера; x, y — значения координат точки эксперта; k_1, k_2 — весовые коэффициенты.

Для каждого эксперта введем понятие нормы как взвешенную метрику (13), в которой вместо центра i-го кластера (x_i,y_i) используется точка (1,1). Будем считать, что чем ближе к точке (1,1) находится эксперт, тем лучше качество его проверки. В этом случае норма (r) может быть вычислена по формуле:

$$r = \sqrt{k_1(1 - x)^2 + k_2(1 - y)^2}\ , \qquad (14)$$

где x, y — значения координат точки эксперта; k_1, k_2 — весовые коэффициенты.

В качестве условия завершения алгоритма рассматривается малое суммарное изменение параметра $dist$ между соседними итерациями среди всех экспертов:

$$\sum dist^{j} - \sum dist^{j-1} < \varepsilon\,, \qquad (15)$$

где ε — заданная константа.

Алгоритм *FCM* чувствителен к резко отклоняющимся значениям оцениваемых параметров — так называемым шумам или выбросам [9]. Поэтому перед применением основного алгоритма производится поиск и исключение таких

точек из исходных данных. Выбросами считаются точки, которые удалены от центра группы на расстояние, превышающее удвоенное среднеквадратическое отклонение этого расстояния для точек данной группы. Начальные центры кластеров распределяются равномерно по центрам отрезков от минимального значения нормированного параметра до максимального. Центры кластеров выбираются таким образом, чтобы минимизировать суммарное расстояние от объектов группы до ее центра и максимизировать расстояние между центрами кластеров. Количество кластеров принимается равным 3 — по количеству основных групп оценивания экспертов (низкое качество проверки, среднее качество проверки и высокое качество проверки).

Начальное значение центра i-го ($i \in [1{:}count]$) кластера ($centr_i$) находится по формуле:

$$centr_i = \min(npar) + i \times \frac{\max(npar) - \min(npar)}{2,0 \times count}, \tag{16}$$

где min (*npar*) — минимальное нормированное значение параметра; max (*npar*) — максимальное нормированное значение параметра; *count* — количество кластеров.

Параметр принадлежности ($belong_i$) эксперта к i-му кластеру находится по формуле [8]:

$$belong_i = \frac{1,0^{\frac{2,0}{n-1,0}}}{dist_i}, \tag{17}$$

где $dist_i$ — дистанция до центра i-го кластера по формуле (13); n — коэффициент нечеткости.

Анализ полученных результатов. Рассмотрим работу предложенного метода на примере обработки результатов проверки экспертной комиссией бланков заданий по дисциплине «Математика: профильный уровень». Комиссия состоит из 59 экспертов, количество проверенных заданий варьируется от 520 до 1462 и в среднем составляет 758.

На рис. 1 и 2 представлены графики распределения экспертов по группам с учетом шумов и без учета шумов соответственно. Эксперты с низким качеством проверки обозначены квадратами, со средним — треугольниками, с высоким — ромбами. Перечеркнутыми фигурами обозначены значения параметров экспертов, которые были приняты выбросами (шумами). Темными точками на графиках обозначены центры кластеров.

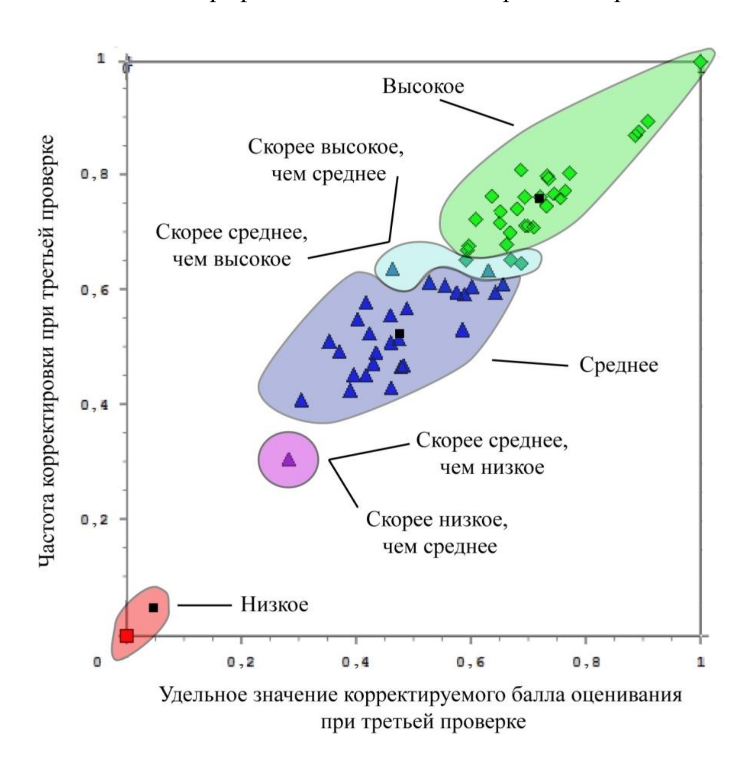


Рис. 1. График распределения экспертов на группы без учета шумов



Рис. 2. График распределения экспертов на группы с учетом шумов

Как видно из рис. 1 и 2, шум оказывает влияние на работу алгоритма при нахождении центров кластеров и является причиной неточного определения качественных параметров экспертов. Неточность определения связана с тем, что шумовые значения могут полностью перетягивать на себя центры кластеров.

Оценка результатов работы комиссии приведена в табл. 3.

Таблица 3

Оценка работы экспертов комиссии с учетом шумов

№	ИН*	3ПрЧастКор, %	3ПрУдВелКор	Выс.	Сред.	Низ.	Результирующая оценка	Шум
1	1506	0,17	0,001718	1,0	0,0	0,0	Высокое	+
2	1211	0,49	0,004902	1,0	0,0	0,0	Высокое	+
3	196	0,55	0,005464	1,0	0,0	0,0	Высокое	+
4	1426	0,57	0,005671	1,0	0,0	0,0	Высокое	–
5	166	0,77	0,009615	1,0	0,0	0,0	Высокое	–
…								
24	229	1,15	0,013436	0,37	0,63	0,0	Скорее среднее, чем высокое	–
25	185	1,16	0,015732	0,2	0,80	0,0	Скорее среднее, чем высокое	–
26	292	1,19	0,01581	0,09	0,91	0,0	Среднее	–
27	632	1,24	0,013201	0,03	0,97	0,0	Среднее	–
28	1505	1,25	0,012545	0,02	0,98	0,0	Среднее	–
…								
38	1427	1,42	0,015986	0,0	0,99	0,01	Среднее	–
39	1425	1,46	0,021938	0,0	0,85	0,15	Скорее среднее, чем низкое	–
40	1428	1,50	0,019461	0,0	0,81	0,19	Скорее среднее, чем низкое	–
41	1507	1,53	0,020443	0,0	0,51	0,49	Скорее среднее, чем низкое	–
42	799	1,55	0,022453	0,0	0,27	0,73	Скорее низкое, чем среднее	–
43	1230	1,61	0,0161	0,0	0,16	0,84	Скорее низкое, чем среднее	–
44	163	1,63	0,021739	0,0	0,025	0,975	Низкое	–

45	633	1,66	0,019934	0,0	0,007	0,993	Низкое	–
46	289	1,68	0,020438	0,0	0,002	0,998	Низкое	–
…								
55	646	1,92	0,020408	0,0	0,0	1,0	Низкое	–
56	638	1,94	0,022887	0,0	0,0	1,0	Низкое	–
57	215	1,99	0,025845	0,0	0,0	1,0	Низкое	+
58	245	2,30	0,026596	0,0	0,0	1,0	Низкое	+
59	1214	3,26	0,036399	0,0	0,0	1,0	Низкое	+
*ИН — идентификационный номер эксперта.								

Рассмотрим пример распределения экспертов на группы с учетом шума. В табл. 3 параметры «Выс.», «Сред.», «Низ.» показывают нормированное значение отношения к кластерам высокого, среднего и низкого качества проверки соответственно. При расчетах принадлежности экспертов к кластерам был выбран коэффициент нечеткости (n) 1,5. Параметр «Результирующая оценка» представлен термами базового и расширенного терм-множества. Эксперты упорядочены по значению их норм (14).

Как видно из табл. 3, показатели шестерых экспертов (1506, 1211, 196, 215, 245, 1214) были приняты за шумы и исключены из алгоритма нахождения центров кластеров. Наихудшие показатели у экспертов 646, 638, 215, 245 и 1214, которые были отнесены к кластеру низкого качества проверки. Некоторые группы экспертов находятся на границах кластеров: высокого и среднего качества проверки (229, 185) и среднего и низкого качества проверки (1425, 1428, 1507, 799, 1230). Наилучшее качество проверки показали эксперты: 1506, 1211, 196, 1426 и 166, которые были отнесены к кластеру высокого качества проверки. У данных экспертов наименьшие показатели частоты корректировок при третьей проверке и удельного корректируемого балла.

Выводы. Проведенный анализ методических рекомендаций ФИПИ по проверке экзамена в форме ЕГЭ показал, что качество работы экспертов отражают следующие параметры: частота корректировки при третьей проверке, удельное значение корректируемого балла оценивания при третьей проверке, среднее значение балла корректировки и среднеквадратическое отклонение балла корректировки.

В результате проведенных экспериментов был выбран двумерный критерий оценки качества работы экспертов, с использованием которого реализовано нечеткое разделение экспертов на классы (высокое, среднее и низкое качество проверки). Оставшиеся показатели могут быть использованы для уточнения тенденции эксперта и определения его уверенности в выставлении баллов.

Разработанная методика позволяет оценить качество работы экспертов и обосновать решения о повышении квалификации членов предметной комиссии.

В соответствии с полученным распределением предлагается при формировании комиссий отказываться от экспертов, которые попали в группу низкого качества проверки. Экспертам со средним качеством проверки следует дать рекомендации, позволяющие улучшить показатели.

Библиографический список

1. Гранков, М. В. Оценка параметров работы экспертов предметной комиссии по проверке экзамена в форме ЕГЭ / М. В. Гранков, Т. А. Щербинин // Системный анализ, управление и обработка информации : тр. 5-го Междунар. семинара / под общ. ред. Р. А. Нейдорфа. — Ростов-на-Дону : Изд. центр ДГТУ, 2014. — С. 282–285.

2. Щербинин, Т. А. Оценка качества работы экспертов, проверяющих экзамены в форме ЕГЭ, с помощью системы кластеризации / Т. А. Щербинин, М. В. Гранков // Автоматизация технологических объектов и процессов. Поиск молодых : сб. науч. работ XV науч.-техн. конф. аспирантов и студентов. — Донецк : ДонНТУ, 2015. — С. 208–211.

3. Гмурман, В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. — Москва : Высшая школа, 2003. — 479 с.

4. Zadeh, L.-A. The Concept of a Linguistic Variable and its Application to Approximate Reasoning-I / L.-A. Zadeh // Information Sciences. — 1975. —Vol. 8. — P. 199–249.

5. Zadeh, L.-A. The Concept of a Linguistic Variable and its Application to Approximate Reasoning-III / L.-A. Zadeh // Information Sciences. — 1975. — Vol. 9. — P. 43–80.

6. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition / F. Hoppner [et al.]. — New York : John Wiley & Sons Inc. — 1999. — 300 p.

7. Kaushik, K. Extension of the fuzzy c means clustering algorithm to fit with the composite graph model for web document representation / K. Kaushik, K. Hemanta // International Journal of Cognitive Research in science, engineering and education. —2013. — Vol. 1, № 2. — 8 p.

8. Рутковский, Л. Методы и технологии искусственного интеллекта / Л. Рутковский. — Москва : Горячая Линия — Телеком, 2010. — 520 с.

9. Анализ данных и процессов / А. А. Баргесян [и др.]. — 3-е изд., перераб. и доп. — Санкт-Петербург : БХВ-Петербург, 2009. — 512 с.

10. Щербинин, Т. А. The method of assessing the quality of work of the Commission to verify exams / Т. А. Щербинин, М. В. Гранков, Е. В. Краснова // Стратегии устойчивого развития мировой науки : тр. V Междунар. науч. конф. Евразийского научного объединения. — Москва : ЕНО, 2015. — С. 31–33.

References

1. Grankov, M.V., Shcherbinin, T.A. Otsenka parametrov raboty ekspertov predmetnoy komissii po proverke ekzamena v forme EGE. [Parameter estimation of experts' work in the subject commission on the assessment of the exam in the form of USE.] Sistemnyy analiz, upravlenie i obrabotka informatsii: Trudy 5 Mezhdunarodnogo seminara. [System analysis, management and information processing: Proc. 5th Int. workshop.] Neydorf, R.A., ed. Rostov-on-Don: DSTU Publ. Centre, 2014, pp. 282–285 (in Russian).

2. Shcherbinin, T.A., Grankov, M.V. Otsenka kachestva raboty ekspertov, proveryayushchikh ekzameny v forme EGE, s pomoshch'yu sistemy klasterizatsii. [Assessment of the quality of experts checking examination in the form of in the form of USE through clustering system.] Avtomatizatsiya tekhnologicheskikh ob"ektov i protsessov. Poisk molodykh. Sbornik nauchnykh rabot KhV nauchno-tekhnicheskoy konferentsii aspirantov i studentov. [Automation of technological objects and processes. Search of the Young: Coll. Sci. papers XV Sci.-Tech. Conf. of graduate and undergraduate students.]Donetsk: DonNTU, 2015, pp. 208–211 (in Russian).

3. Gmurman, V.E. Teoriya veroyatnostey i matematicheskaya statistika. [Theory of probability and Mathematical statistics.] Moscow: Vysshaya shkola, 2003, 479 p. (in Russian).

4. Zadeh, L.-A. The Concept of a Linguistic Variable and its Application to Approximate Reasoning-I. Information Sciences, 1975, v*ol.* 8, pp. 199–249.

5. Zadeh, L.-A. The Concept of a Linguistic Variable and its Application to Approximate Reasoning-III. Information Sciences, 1975, v*ol.* 9, pp. 43–80.

6. Hoppner, F., et al. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. New York: John Wiley & Sons Inc., 1999, 300 p.

7. Kaushik, K., Hemanta, K. Extension of the fuzzy c means clustering algorithm to fit with the composite graph model for web document representation. International Journal of Cognitive Research in science, engineering and education, 2013, vol. 1, no. 2, p. 8.

8. Rutkovskiy, L. Metody i tekhnologii iskusstvennogo intellekta. [Methods and technologies of artificial intelligence.] Moscow: Goryachaya Liniya — Telekom, 2010, 520 p. (in Russian).

9. Bargesyan, A.A., et al. Analiz dannykh i protsessov. [Analysis of data and processes.] 3rd revised and enlarged ed. SPb.: BVKh-Peterburg, 2009, 512 p. (in Russian).

10. Shcherbinin, T.A., Grankov, M.V., Krasnova, E.V. The method of assessing the quality of work of the Commission to verify exams. Strategies for sustainable development of the world science: Proc. V Int. Sci. Conf. Eurasian scientific association. Moscow: ENO, 2015, pp. 31–33.

Поступила в редакцию 01.02.2016
Сдана в редакцию 01.02.2016
Запланирована в номер 07.07.2016